

# GENERATING REPRESENTATIVE SEQUENCES OF DAILY PRECIPITATION FOR AGRICULTURAL SIMULATIONS

J. D. Garbrecht, J. X. Zhang

**ABSTRACT.** *Generated weather that represents alternative realizations of a particular historical record is often needed for computer simulation of agricultural or agronomic impacts. The stochastic component of generated weather is based on random numbers. This study shows that relatively short sequences of uniform random numbers, as used to generate daily precipitation, may not display sufficient uniformity to reproduce the distribution of the target historical precipitation record or of a seasonal precipitation forecast. The magnitude of the discrepancy becomes an issue in applications that require accurate replication of the historical record or forecast. A procedure is proposed to test sequences of uniform random numbers before their use and retain only those sequences that display a high degree of uniformity as required by the weather generation model. Generated precipitation with and without testing for uniformity of random numbers shows that the discrepancy between the two can be important particularly for simulation duration less than 50 years, as are often involved in practical water resources and agricultural applications. The use of "tested" random numbers leads to transition probabilities and a precipitation distribution that are generally closer in agreement with those of the historical record. The effectiveness of the method is illustrated by comparing generated daily precipitation based on "tested" and "non-tested" random number for the Kingfisher, Oklahoma, historical precipitation record. The increased representativeness of the generated precipitation using "tested" random numbers allows for shorter simulation duration of agricultural models and greater capability to simulate subtle changes in precipitation such as those associated with seasonal precipitation forecasts.*

**Keywords.** *Precipitation-generation, Computer-simulation, Precipitation, Weather, Climate.*

Stochastically generated daily weather is often used to drive environmental and agricultural models to simulate erosion, chemical transport, productivity, or management issues (e.g. Richardson, 1981; Williams et al., 1984; Richardson, 1985; Wilks, 1992; Woo, 1992; Katz, 1996; Mearns et al., 1996; Mavromatis and Hansen, 2001). Generated weather is needed to supplement existing weather data, provide alternative weather realizations for a particular historical record, or identify possible weather sequences for a seasonal climate forecast. Previous studies have generally focused on model development and validation of weather generators (e.g. Richardson, 1982; Wallis, 1993; Johnson et al., 1996; Semenov et al., 1998; Wilks, 1999; Wilks and Wilby, 1999). Here practical consideration is given to situations where it is essential that the generated daily precipitation reproduce closely the distribution or summary statistics of a historical record or a seasonal forecast.

One generally expects and assumes that generated weather accurately reproduce the summary statistics that are

used to drive the weather generator. Yet recalculation of mean transitional probabilities and mean daily precipitation from the generated daily precipitation shows that this is not always the case due to the stochastic nature of generated weather. The average size of the discrepancy depends largely on the length of the generated weather sequence. The longer the generated weather sequence, the closer its summary statistics were to the expected values. The typical length of generated weather sequences for agricultural analyses is between 30 and 50 years. For such short lengths the discrepancy in summary statistics of the generated precipitation can be of the same order of magnitude as departures associated with a typical seasonal precipitation forecast. Hence, discrepancies related to the stochastic nature of generated precipitation may overshadow the intended simulation of a systematic difference associated with a particular seasonal precipitation forecast. Theoretically, this shortcoming can be overcome by simply increasing the length of the sequence of generated precipitation to several hundreds of years. However, considerations related to subsequent use of the generated weather in complex and resource intensive agricultural simulation models suggest that sequences of about 50 years are more prevalent in practice. Hence, a procedure for generating daily precipitation sequences that consistently and more closely reproduce the underlying summary statistics within about 50 years of generated weather is needed to address such special applications.

In this study, a procedure is proposed to generate 50- to 100-year long sequences of daily precipitation that closely approximate target summary statistics representing either a historical record or a seasonal climate forecast. Specifically, fundamental assumptions underlying the daily precipitation

---

Article was submitted for review in October 2002; approved for publication by the Soil & Water Division of ASAE in May 2003.

Names of precipitation models and random number generators are given for illustration purposes only, and no comparison, rating, or endorsement is intended.

The authors are **Jurgen D. Garbrecht**, Research Hydraulic Engineer, and **John X. Zhang**, Hydrologist, USDA-ARS, Grazinglands Research Laboratory, El Reno, Oklahoma. **Corresponding author:** Jurgen Garbrecht, USDA-ARS, 7207 W. Cheyenne, El Reno, OK 73036; phone: 405-262-5291; fax: 405-262-0133; e-mail: garbrech@grl.ars.usda.gov.

models are reviewed, and properties of uniform random number sequences relevant to precipitation generation are discussed. The proposed adaptation of the precipitation model is presented, and the improved capabilities are illustrated by a practical application to the historical record from the Kingfisher, Oklahoma, weather station. The proposed procedure calls for a strict uniform distribution of the sequence of uniform random numbers used to generate precipitation values. This approach ensures a high level of compatibility between model assumptions of a uniform distribution and uniformity of the actual sequence of random number that is used. The increased compatibility between model assumptions and random number properties leads to a better replication of specific target statistics by reducing unintended random variability. The use of so called “tested” random numbers in the precipitation generation model is particularly useful when a rapid convergence to target summary statistics is desired, and for practical environmental and agricultural model applications that simulate subtle changes in weather characteristics.

## STOCHASTIC GENERATION OF DAILY PRECIPITATION

### MODEL DESCRIPTION

Weather generation computer programs, such as CLIGEN (Nicks and Gander, 1994), WGEN (Richardson and Wright, 1984), US CLIMATE (Hanson et al., 1994), and GEM (Johnson et al., 2000), often use a chain-dependent stochastic precipitation model to characterize daily precipitation at a site. A chain-dependent stochastic precipitation model is also used in this investigation. Daily precipitation is modeled in two sequential parts: first, the occurrence of a rainy day, and second, the amount of precipitation on a rainy day (Waymire and Gupta, 1981). A two-state, first-order Markov chain determines the occurrence of rainy days. The Markov chain is said to be two-state because it only considers whether precipitation does or does not occur on a specific day, and it is first-order because the probability of precipitation on a rainy day is conditioned only on the precipitation state of the previous day (Haan et al., 1976). The probabilities of precipitation given the previous day’s precipitation state are called transition probabilities. The transition probabilities considered here are the probability of a wet day after a dry day (PWD), and of a wet day after a wet day (PWW). The nonstationarity of transition probabilities within a year due to seasons is introduced by using separate transition probability values for each calendar month (Wilks, 1989). The precipitation amount on a rainy day is specified by a probability distribution function (or cumulative distribution function) of daily precipitation amounts. In this investigation a double exponential distribution is used. As before, seasonality is introduced by using different distribution parameters for each calendar month. Both the transition probabilities and distribution parameters are derived from available historical data. Further information on the fundamentals of chain-dependent stochastic precipitation models can be found in Woolhiser and Roldan (1982), Hanson et al. (1994), Johnson et al. (1996), and Parlange and Katz (2000).

### STOCHASTIC PRECIPITATION GENERATION PROCESS

Occurrence of a rainy day is determined by comparing a random number generated from a uniform distribution between 0 and 1 to the value of the transition probabilities PWD or PWW. If the preceding day is dry and the random number is smaller than PWD, then the current day is a rainy day; alternatively, if the random number is greater than PWD, then the current day is dry. The decision process is similar if the preceding day is wet. Once the occurrence of a rainy day has been established, the amount of precipitation on that day is determined by generating a new random number from a uniform distribution and solving the inverse cumulative distribution function for daily precipitation, i.e. the random number is taken as the cumulative frequency value and the corresponding daily precipitation is determined analytically or numerically (fig. 1). Strings of consecutive random numbers (called sequences) are generated separately for each transition probability, precipitation distribution, and calendar month. For 30- or 50-year’s worth of generated precipitation data, anywhere between 50 to 1500 random numbers are required per parameter and calendar month. Generation of separate sequences of random numbers for each parameter and month is important, because selectively picking random numbers out of a single continuous string for assignment to different parameters and time periods may alter the random properties of the resulting sequence of random numbers for that parameter and time period. For example, generation of 30 years of daily precipitation for the month of January (assuming, on average, 29 dry days and 2 rainy days) requires a sequence of about 900 uniformly distributed random numbers (30 years  $\times$  29 dry days) to determine the occurrence of rainy days after dry days. To determine the amount of precipitation on each of the rainy days requires a separate sequence of 60 uniformly distributed random numbers (30 years  $\times$  2 rainy days). And to determine the occurrence of rainy days after wet days, requires another separate sequence of 60 uniformly distributed random numbers (30 years  $\times$  2 wet days). In arid climates, the scarcity of rainy days may require that over 50 years of precipitation be generated to adequately reproduce the historical record.

### PRECIPITATION MODEL ASSUMPTIONS AND LIMITATIONS

First, precipitation is assumed to be stationary in time. Thus, systematic annual variations or trends are not modeled. Second, transition probabilities and daily precipitation distributions are assumed to be monthly independent. Thus, non-random persistence effects of monthly precipitation [i.e. as a result of El Nino Southern Oscillation or other

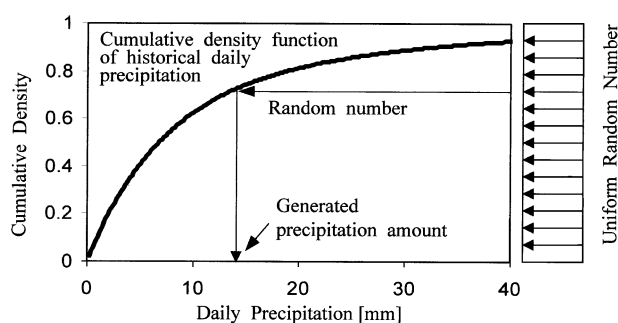


Figure 1. Schematic of the generation of daily precipitation amounts.

ocean–atmosphere anomalies (Woolhiser et al., 1993)] are not modeled. Third, random numbers used for precipitation generation are assumed to be uniformly distributed between 0 and 1. Anything other than uniformly distributed random numbers will not reproduce or converge to the transition probabilities and daily precipitation distribution of the historical data that is being simulated. Fourth, daily precipitation is generally modeled by medium–tailed distributions (exponential, gamma, mixed–exponential, skewed–normal, etc.). Thus, one should not expect the model to work well for the generation of extreme daily precipitation values that have a heavy–tailed distribution (Katz, 2002). Fifth, monthly transition probabilities and daily precipitation distributions are derived unconditionally, i.e. they are not conditioned on wet or dry months. As a result low– and high–end monthly precipitation values are not well reproduced (Wilks, 1989). In the context of these assumptions and limitations, the role and effect of random numbers on generated daily precipitation is investigated.

## RANDOM NUMBERS

### RANDOM NUMBER GENERATORS

Random numbers generated from three different random number generators are used to graphically illustrate typical variations in random number sequences. The first random number generator, RN1, is by MacLaren and Marsaglia (1965) and was used in the original version of CLIGEN. The second generator, RN2, is by Press et al. (1992) and includes two generators and shuffling to break up any sequential correlations. And, the third generator, RN3, is a library function included in the Salford compiler (Salford, 1998).

### RANDOM NUMBER PATTERNS

Thirty thousand (30,000) random numbers were generated with each of the three random number generators. Localized systematic departures of sequences of random numbers from the expected mean value are visualized by plotting, for each generator, the cumulative departures of the random numbers from the expected mean of 0.5 (fig. 2). Cumulative departures are defined as:

$$cd_n = \sum_{i=1}^n (x_i - 0.5) \quad (1)$$

where  $cd$  is cumulative departure,  $n$  is the ending position in the sequence of random numbers, and  $x_i$  is the uniform random numbers. If sequences of random numbers were uniformly distributed between 0 and 1, then the expected value of the cumulative departure should be zero and the curve of cumulative departures should randomly fluctuates about the zero line. However, local steep and sustained positive and negative variations in the plotted data (indicated by A in fig. 2) are observed. Such variations are indicative of sequences of random numbers that display, over the length of the sequence (up to 1000 values), systematic departures from the mean of the uniform distribution. Steeper variations are indicative of larger departures. The departures are most likely due to inherent properties of sequences of random number (random clustering effects). The observed departures of short sequences of random numbers within the overall set of random numbers do not imply that the random numbers are

not random. It merely points to the fact that short sequences of random numbers may not closely adhere to the uniform distribution due to the stochastic nature of random numbers. Hence, while the generated random numbers display correct random properties, short sequences of random numbers may depart from the expectation of a uniform distribution. Thus, if such a sequence of random numbers were to be used for precipitation generation, then the generated precipitation may not be representative of the intended historical data. For example, in figure 1, if the distribution of the random numbers were not uniform but trapezoidal with more values in the lower portion, then a larger number of smaller daily precipitation values would be generated. The practical implications of non–uniformly distributed random numbers are illustrated in the example application section.

### DEVIATION OF RANDOM NUMBERS FROM EXPECTED VALUES

For each curve in figure 2, the mean of non–overlapping sequences of 50, 100, and 200 random numbers is calculated. The length of sequences corresponds approximately to the minimum number required for 30– to 50–year’s worth of precipitation generation for a particular parameter and calendar month. The mean deviation, standard deviation, and maximum deviation are used as measures of departure and are expressed in percent of the expected mean (0.5). Deviation refers to the absolute difference between the mean of a generated sequence of random numbers and the expected mean of 0.5. The results are summarized in table 1. The magnitude of the deviation is similar for the three generators. The values in table 1 are in line with what one would expect based on random sampling from a uniform distribution. For example, the standard deviation of the mean (in percent of mean) for sequences of 50, 100, and 200 generated random numbers is 8.3%, 5.8%, and 3.9%, respectively, which is close to the theoretically expected values of 8.16%, 5.77%, and 4.08%, respectively. The average of the mean deviation (last column in table 1) for non–overlapping sequences of 100 random numbers is 4.6%, the average standard deviation is 5.8%, and the average maximum deviation is 18.0%. The same statistics for sequences of 50 random numbers (first three rows in table 1) and 200 random numbers (last three rows in table 1) show that quadrupling the sequence length from 50 to 200 random numbers reduces the average

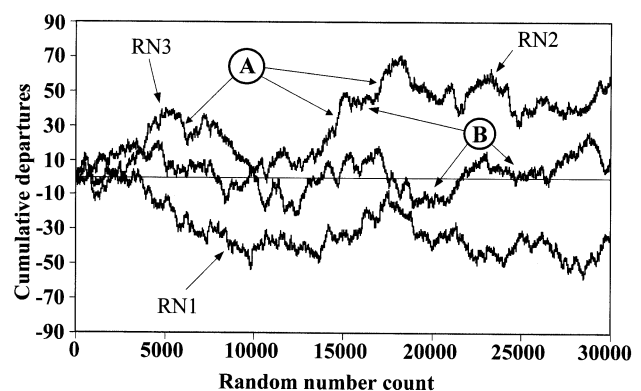


Figure 2. Cumulative departures of 30,000 random numbers generated by three different random number generators (RN1, RN2, RN3). Symbol A shows locations where the sequence of random numbers is not likely uniformly distributed, and symbol B shows locations where it is likely uniformly distributed.

**Table 1. Deviations of the mean from three random number generators of non-overlapping sequences of 50, 100, and 200 random numbers from the expected value of 0.5.**

Generator Name	RN1 (%)	RN2 (%)	RN3 (%)	Average (%)
600 sequences of 50 random numbers				
Mean deviation	6.5	6.9	6.5	6.6
Standard deviation	8.0	8.6	8.2	8.3
Maximum deviation	24.5	32.7	27.7	28.3
300 sequences of 100 random numbers				
Mean deviation	4.8	4.6	4.4	4.6
Standard deviation	6.1	5.8	5.5	5.8
Maximum deviation	20.2	18.7	15.2	18.0
150 sequences of 200 random numbers				
Mean deviation	2.9	3.2	3.3	3.1
Standard deviation	3.5	4.0	4.3	3.9
Maximum deviation	9.9	11.5	15.3	12.2

deviation by only about half, which is what one should expect with increasing sample size.

Given the pattern of localized systematic departures (fig. 2), the size of the deviations (table 1), and the requirement that random numbers conform to the uniform distribution (precipitation model), one could be inclined to test sequences of random numbers for conformity to the uniform distribution and only use those sequences that pass that test. Such an approach has been proposed by Niederreiter (1978) who argued that statistical randomness may not be the only desirable property, and that other properties of generated random numbers, such as “evenness” of the distribution of random points, may also be important in some applications. This is believed to be the case here where the generated precipitation can only reproduce the desired historical precipitation distribution if the random numbers are truly uniformly distributed over the range of 0 and 1. The notion of making certain properties of a sequence of random numbers dependent on the intended use has also been suggested in Law and Kelton (2000).

#### RANDOM NUMBER TESTING

Law and Kelton (2000) proposed to use a Chi-Square “goodness-of-fit” test, with all parameters known, to check whether a sequence of random numbers is uniformly distributed between 0 and 1. In this study the authors tested sequences of 100 random numbers as they are generated. The value of 100 random numbers was a compromise between the low-end value for the generation of 30 to 50 years’ worth of precipitation and the need for large samples to conduct a meaningful Chi-Square test. Traditionally, one would reject random numbers that fail the test with a  $\alpha$  value of 0.1 or 0.05, meaning that only random numbers that have a truly uneven distribution are rejected. Here, however, we want to retain random numbers that are very evenly distributed, and, therefore, the test is conducted with a  $\alpha$  value of 0.95. In addition to uniformity of the distribution, the mean of the random numbers is also tested to be within 0.5% or less of the expected mean of 0.5. Hence, sequences of 100 random numbers that pass these tests can be expected to approximate the uniform distribution very closely. These sequences of random numbers are referred to as “tested” random numbers because they are made to conform to the uniformity assumption of the precipitation model. In the practical

implementation of this test, a suitable sequence of random numbers cannot always be found within a desired number of tries. In these instances the Chi-Square test criteria are relaxed, i.e.  $\alpha$  is incrementally reduced (0.94, 0.93, 0.92, 0.91, and 0.90), and the difference in the mean is incrementally increased (1.0%, 1.5%, and 2%), until a pool of random numbers that passes the test is found.

Thirty thousand “tested” random numbers were generated and non-overlapping sequences of 50, 100, and 200 random numbers were evaluated in an identical manner as the “non-tested” random numbers in the previous section. Results are shown in table 2. Mean deviation, standard deviation, and maximum deviation are all smaller than corresponding values for the three random number generators presented in the previous section (table 1). For sequences of 50 random numbers, corresponding values decreased by about 30%, whereas for sequences of 100 and 200 random numbers values decreased by a factor of about 10 or larger. Sequences of 100 and 200 random numbers fared much better because the testing for uniformity was conducted on pools of 100 random numbers. Overall, “tested” random numbers adhere much closer to the assumption of uniform distribution than “non-tested” random numbers. Similar high levels of improvements could have been achieved for the sequence of 50 random numbers if the testing were conducted on pools of 50 random numbers. The underlying assumption here is that any practical application of the weather generator will use at least 100 random numbers per variable and month, and testing can be conducted on pools of 100 random numbers.

#### IMPACT ON GENERATED PRECIPITATION

The daily precipitation data of the National Weather Service station at Kingfisher, Oklahoma is used as baseline to illustrate the impact of “tested” versus “non-tested” random numbers on the generated daily precipitation. The impact is defined in terms of deviation of generated daily precipitation statistics from baseline values of the historical precipitation data, expressed in percentage of the historical values. Deviations are calculated by month for the transition probabilities PWD and PWW and for the mean, standard deviation and skew coefficient of the daily precipitation. The impact of “tested” versus “non-tested” random numbers on generated precipitation is evaluated for ten separate precipitation generation runs of 50 years each. Ten separate generation runs are made, each with different random numbers. The different random numbers result in slightly different generated precipitation values. The random number generator used for this assessment is RN3. This random number generator was selected because it is the most detailed of the three used in this study.

**Table 2. Deviations of the mean for non-overlapping sequences of 50, 100, and 200 “tested” random numbers, expressed in percent of the expected value of 0.5.**

	600 Sequences of 50 Random Numbers (%)	300 Sequences of 100 Random Numbers (%)	150 Sequences of 200 Random Numbers (%)
Mean deviation	4.8	0.4	0.3
Standard deviation	6.0	0.5	0.4
Maximum deviation	18.8	1.4	0.9

**Table 3. Deviations of transition probabilities PWD (rainy day after dry day) and PWW (rainy day after rainy day) from historical baseline data in percent of baseline value.**

	Jan.	Feb.	March	April	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.	Avg.
PWD													
Non-tested	6.5	5.5	5.8	6.0	3.5	4.9	5.8	4.6	4.3	5.1	6.1	5.5	5.3
Tested	3.4	3.0	3.5	2.3	2.7	2.4	3.8	3.3	2.2	2.1	3.7	3.9	3.0
PWW													
Non-tested	11.3	8.2	7.8	4.1	5.3	4.1	6.7	4.4	6.8	7.1	9.6	8.5	7.0
Tested	6.6	2.6	4.1	2.6	1.3	2.7	4.6	2.9	2.6	1.9	1.6	5.2	3.2

### TRANSITION PROBABILITIES

Transition probabilities PWD and PWW are calculated for the ten generated precipitation data sets, and average monthly deviations from the historical baseline are reported in table 3 for the case of “non-tested” and “tested” random numbers. For all months the deviation of transition probabilities from baseline is smaller for “tested” random numbers. The largest average deviation in PWD produced by “non-tested” random numbers is 6.5% (Jan.), whereas it is 3.9% (Dec.) for “tested” random numbers. Similarly, for PWW, the values are 11.3% and 6.6%, respectively. On average over all months (last column in table 3), the deviations for PWD and PWW are reduced by a factor of two when using “tested” versus “non-tested” random numbers.

### DAILY PRECIPITATION DISTRIBUTION

The deviations in mean, standard deviation, and skew coefficient of the generated daily precipitation are presented in table 4 for each month and as an average over the year. Deviations in the mean (first two rows) appear to be systematically smaller for “tested” than for “non-tested” random numbers. The annual deviation (last column in table 4) shows that the average reduction is about a factor of two. However, for this application, the reduction does not extend to the standard deviation and skew coefficient that have similar deviations for “tested” and “non-tested” random numbers.

The distributions of generated daily precipitation for the month of April for each of the 10 generation runs are plotted in figure 3. Each distribution of generated daily precipitation attempts to reproduce the fitted three-parameter, double-exponential distribution of the historical data at Kingfisher, Oklahoma (thick line in fig. 3). The distributions resulting from “non-tested” random numbers (fig. 3 top) show a much greater spread than those distributions resulting from “tested” random numbers (fig. 3 bottom). At the median, the range of the probability of exceedance curves of daily precipitation is 0.9 mm for the “tested” random numbers, down from 2.5 mm for “non-tested” random numbers. The greater spread associated with “non-tested” random numbers

is attributed to the sustained departures of sequences of random numbers from the uniform distribution as shown in figure 2. The results show that daily precipitation values that have been generated with “tested” random numbers approximate the fitted distribution of the historical data at Kingfisher better and more consistently.

### DISCUSSION

Mean monthly precipitation derived from daily precipitation generated based on “tested” random numbers is well reproduced; yet the variability of the derived monthly precipitation is often underestimated. A similar underestimation of variability was also reported for the case where “non-tested” random numbers are used in the model (Johnson et al., 1996; Wilks, 1989). A fundamental reason for this underestimation of monthly variability lies in the simplifying assumptions of the precipitation model, and not so much in the quality and properties of the random numbers. The model assumes constant monthly transition probabilities and daily precipitation distributions. In reality, occurrence of rainy days and amount of daily precipitation change when a particular month is either dry or wet (Wilks, 1989). Wilks showed that daily precipitation generation conditioned on dry/wet state of a month will reproduce the high and low end of monthly precipitation. Based on the similarity in the average departure of the standard deviation (table 4) between daily precipitation generated from “non-tested” and “tested” random numbers, it is believed that the use of “tested” random numbers for precipitation generation does not contribute much to degrading the variability of monthly precipitation beyond that already inherent to model simplifications.

Similar considerations apply for variations in annual precipitation that often are a function of persistence of wet or dry months. The daily precipitation model assumes independence between monthly precipitation statistics and, thus, allows only simulation of random, probabilistic clustering of wet or dry months (Johnson et al., 1996). This limits

**Table 4. Deviations of the mean, standard deviation, and skew coefficient of generated daily precipitation from historical baseline data in percent of baseline value.**

	Jan.	Feb.	March	April	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.	Avg.
Mean													
Non-tested	7.2	5.7	5.7	7.2	4.8	3.6	6.3	6.1	6.2	6.7	8.3	6.0	6.2
Tested	3.4	6.1	2.5	1.3	2.1	2.8	2.6	3.7	2.3	2.9	4.0	2.6	3.0
St. Dev.													
Non-tested	9.5	8.1	11.2	8.5	5.4	3.2	4.9	8.6	7.2	11.8	8.9	8.1	7.9
Tested	7.2	18.7	7.7	2.9	3.9	5.4	7.8	9.0	7.4	5.4	7.1	6.9	7.4
Skew Coef.													
Non-tested	28.8	21.5	29.6	11.3	10.2	7.9	16.5	17.4	14.9	18.3	23.4	15.2	17.9
Tested	20.0	24.9	25.6	14.1	11.8	11.1	27.1	15.9	19.0	17.2	24.0	17.7	19.0

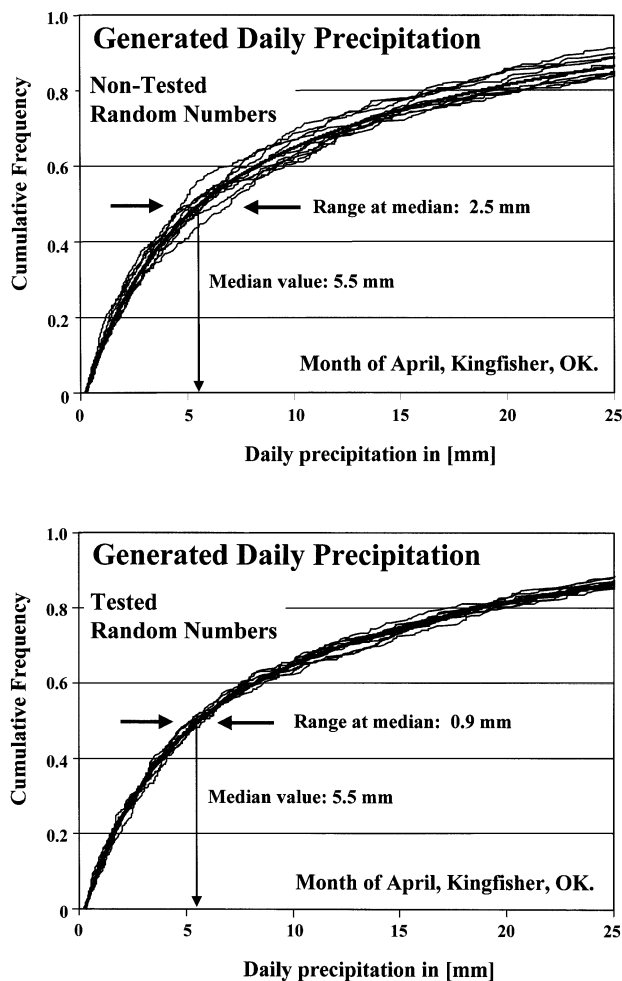


Figure 3. Distributions of generated daily precipitation based on “non-tested” (top) and “tested” (bottom) random numbers. The thick line at the center of the generated distributions represents the fitted distribution of the historical data at Kingfisher weather station, Oklahoma.

the model’s ability to simulate systematic, long-term wet/dry persistence, such as those brought about by an El Nino or other low frequency ocean–atmosphere interactions (Woolhiser et al., 1993). Here again, the underestimation of annual precipitation variability is induced by model simplifications, and should not be attributed to the use of “tested” random number in the precipitation generation process.

One may also be inclined to think that additional variability associated with “non-tested” random number sequences may provide a basis for generating more accurately extreme daily precipitation events of 100– to 200-year length. Such a thought should be viewed with great skepticism. First, there are no reasons to believe that “non-tested” random numbers should occur more frequently at the edge of the distribution (near 0.0 or 1.0) than “tested” random numbers. A test of 100,000 random numbers showed that both the “non-tested” and “tested” approach produced about the same number of values above a threshold of 0.999, with highest value in either approach near 0.99999. Second, generally accepted distributions of daily precipitation (exponential, gamma, mixed-exponential, skewed-normal, etc.)

are medium-tailed distributions and less suitable for accurate representation of extreme values which generally have a heavy-tailed distribution. And, third, the focus and intent of the proposed approach is to produce repeat sequences that approximate the intended summary statistics well, not to generate more accurate extreme events. Given these considerations the use of “tested” random numbers do not negatively impact the generation of extreme events.

The best demonstration that model-generated daily precipitation based on “tested” random number sequences do approximate the fitted distribution of the historical record or a seasonal precipitation forecast (target statistics) more consistently is illustrated in figure 3. The use of “tested” random numbers leads to distributions of generated daily precipitation that are more consistent with one another and with the target distribution. Thus, the proposed approach fulfills the objective of reproducing accurately a desired target distribution within constraints of commonly used simulation durations for agricultural applications.

## CONCLUSIONS

This study recognized that short sequences of uniform random numbers, while being acceptably uniform in the terms of random sampling statistics, do not always meet uniformity expectations for daily precipitation generation that aim at consistently reproducing the summary statistics of a historical record or a seasonal precipitation forecast. For such purpose, the model for daily precipitation generation must be driven by random numbers that are uniformly distributed between 0 and 1. Yet, short sequences of random numbers are not necessarily uniformly distributed, and discrepancies may arise between expected and actual properties of random numbers provided to the precipitation generation model. This limitation is not new and has been well formulated by Neiderreiter (1978): “... instead of trying to cope with the impalpable concept of randomness, one should select [random] points according to a deterministic scheme that is well suited for the problem at hand.” In this study, a deterministic testing procedure has been proposed to screen short sequences of random number to meet the requirements of the precipitation generation model, i.e. random numbers uniformly distributed between 0 and 1. The following conclusions can be drawn from this investigation:

- The use of “tested” versus “non-tested” random numbers in stochastic generation of daily precipitation does make a difference in the characteristics of generated daily precipitation data and can be a factor when close reproducibility of a target distribution is an issue.
- Daily precipitation generated with “tested” random numbers consistently produce transition probabilities and a mean of daily precipitation that better approximate the historical record or a seasonal precipitation forecast, and promoted a faster convergence to the target summary statistics.
- The improvements achieved by use of “tested” sequences of random numbers enhance the usefulness of generated precipitation in practical simulation applications that need to identify agronomic impacts of a particular climate or

forecast departure within constraints of realistic simulation durations.

## REFERENCES

- Haan, C. T., D. M. Allen, and J. O. Street. 1976. A Markov chain model of daily precipitation. *Water Resources Research* 12( 3): 443–449.
- Hanson, C. L., K. A. Cumming, D. A. Woolhiser, and C. W. Richardson. 1994. Microcomputer program for daily weather simulation in the contiguous United States. United States Department of Agriculture, Agricultural Research Service. Washington, D.C.
- Johnson, G. L., C. L. Hanson, S. P. Hardegree, and E. B. Ballard. 1996. Stochastic weather simulation: Overview and analysis of two commonly used models. *J. of Applied Meteorology* 35(10): 1878–1896.
- Johnson, G. L., C. Daly, G. H. Taylor, and C. L. Hanson. 2000. Spatial variability and interpolation of stochastic weather simulation model parameters. *J. of Applied Meteorology* 39(6): 778–796.
- Katz, R. W. 1996. Use of conditional stochastic models to generate climate change scenarios. *Climate Change* 32: 237–255.
- Katz, R. W. 2002. Do weather or climate variables and their impacts have heavy-tailed distributions? In *Proc. of 16th Conference on Probability and Statistics in the Atmospheric Sciences*, J84–J87. Boston, Mass.: American Meteorological Society.
- Law, A. M., and W. D. Kelton. 2000. *Simulation Modeling and Analysis*, 3rd Ed. Boston, Mass.: McGraw Hill Inc.
- MacLaren, D. M., and G. Marsaglia. 1965. Uniform random number generators. *J. of Association for Computing Machinery* 12(1): 83–89.
- Mavromatis, Y., and J. Hansen. 2001. Interannual Variability characteristics and simulated crop response of four stochastic weather generators. *Agricultural and Forest Meteorology* 109: 283–296.
- Mearns, L., C. Rosenzweig, and R. Goldberg. 1996. The effect of changes in daily and interannual climate variability on CERES–wheat: A sensitivity study. *Climate Change* 32: 257–292.
- Nicks, A. D., and G. A. Gander. 1994. CLIGEN: A weather generator for climate inputs to water resources and other models. *Proc. Fifth Int. Conf. on Computers in Agriculture*, 903–909. St. Joseph, Mich.: ASAE.
- Niederreiter, H. 1978. Quasi–Monte Carlo methods and pseudo-random numbers. *Bulletin of the American Mathematical Society* 84(6): 957–1041.
- Parlange, M. B., and R. W. Katz. 2000. An extended version of the Richardson model for simulating daily weather variables. *J. of Applied Meteorology* 39( 5): 610–622.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. *Numerical Recipes in FORTRAN77: The Art of Scientific Computing*, Second Ed. New York: Cambridge University Press.
- Richardson, C. W. 1981. Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research* 17( 1): 182–190.
- Richardson, C. W. 1982. A comparison of three distributions for the generation of daily rainfall amounts. In *Statistical Analysis of Rainfall and Runoff*. Littleton, Colo.: Water Resources Publications.
- Richardson, C. W. 1985. Weather simulation for crop management models. *Transactions of the ASAE* 28(5): 1602–1606.
- Richardson, C. W., and D. A. Wright. 1984. WGEN: a model for generating daily weather variables. USDA Agricultural Research Service Publication ARS–8. [Available from U. S. Dept. of Agriculture, Agricultural Research Service, 808 E. Blackland Rd., Temple, Tex.]
- Salford. 1998. FTN95 Reference Library. Salford, UK: Salford Software Ltd.
- Semenov, M., R. Brooks, E. Barrow, and C. Richardson. 1998. Comparison of the WGEN and LARS–WG stochastic generators for diverse climates. *Climate Research* 10: 95–107.
- Wallis, T. W. R. 1993. A comparison between weather simulated within the erosion/productivity impact calculator (EPIC) and observed data. *Intl. J. of Environmental Studies* 45: 1–12.
- Waymire, E., and V. K. Gupta. 1981. The mathematical structure of rainfall representations 1. A review of the stochastic rainfall models. *Water Resources Research* 7(5): 1261–1272.
- Wilks, D. S. 1989. Conditioning stochastic daily precipitation models on total monthly precipitation. *Water Resources Research* 25(6): 1429–1439.
- Wilks, D. S. 1992. Adapting stochastic weather generation algorithms for climate change studies. *Climate Change* 22: 67–84.
- Wilks, D. S. 1999. Interannual variability and extreme–value characteristics of several daily precipitation models. *Agricultural and Forest Meteorology* 93: 153–169.
- Wilks, D. S., and R. L. Wilby. 1999. The weather generation game: a review of stochastic weather models. *Progress in Physical Geography* 23(3): 329–357.
- Williams, J. R., C. A. Jones, and P. T. Dyke. 1984. A modeling approach to determining the relationship between erosion and soil productivity. *Transactions of the ASAE* 27(1): 129–144.
- Woo, M. 1992. Application of stochastic simulation to climate change studies. *Climate Change* 20: 313–330.
- Woolhiser, D. A., and J. Roldan. 1982. Stochastic daily precipitation models 2. A comparison of distribution amounts. *Water Resources Research* 18(5): 1461–1468.
- Woolhiser, D. A., T. O. Keefer, and K. T. Redmond. 1993. Southern oscillation effects on daily precipitation in the southwestern United States. *Water Resources Research* 29(4): 1287–1295.

## NOMENCLATURE

A	=	significance level of the Chi–Square test
Cd	=	cumulative departure
I	=	counter
N	=	upper count limit for cd calculation
PWD	=	Probability of rainy day after dry day
PWW	=	Probability of rainy day after rainy day
RN1	=	random number generator by MacLaren and Marsaglia (1965)
RN2	=	random number generator by Press et al. (1992)
RN3	=	random number generator of Salford compiler (Salford, 1998)
X	=	random number

